

## RESEARCH ARTICLE

# Crystal Structure of the 20S Proteasome from the Archaeon *T. acidophilum* at 3.4 Å Resolution

Jan Löwe, Daniela Stock, Bing Jap,\* Peter Zwickl,  
Wolfgang Baumeister, Robert Hubert

The three-dimensional structure of the proteasome from the archaeobacterium *Thermoplasma acidophilum* has been elucidated by x-ray crystallographic analysis by means of isomorphous replacement and cyclic averaging. The atomic model was built and refined to a crystallographic *R* factor of 22.1 percent. The 673-kilodalton protease complex consists of 14 copies of two different subunits,  $\alpha$  and  $\beta$ , forming a barrel-shaped structure of four stacked rings. The two inner rings consist of seven  $\beta$  subunits each, and the two outer rings consist of seven  $\alpha$  subunits each. A narrow channel controls access to the three inner compartments. The  $\alpha_7\beta_7\beta_7\alpha_7$  subunit assembly has 72-point group symmetry. The structures of the  $\alpha$  and  $\beta$  subunits are similar, consisting of a core of two antiparallel  $\beta$  sheets that is flanked by  $\alpha$  helices on both sides. The binding of a peptide aldehyde inhibitor marks the active site in the central cavity at the amino termini of the  $\beta$  subunits and suggests a novel proteolytic mechanism.

The proteasome is now widely recognized as the central enzyme of nonlysosomal protein degradation. It is involved in the degradation of misfolded proteins (1) as well as in the degradation and processing of short lived regulatory proteins (2). Peptide aldehyde inhibitors of the proteasome have been shown to block the degradation of most cellular proteins, including long-lived proteins and the generation of peptides presented on major histocompatibility (MHC) class I molecules *in vivo* (3). Thus the life-span of many cellular proteins seems to be controlled by this proteolytic system. These 20S proteasomes occur in the cytosol and in the nuclei of all eukaryotic cells examined so far, and they have also been found in the archaeobacterium *Thermoplasma acidophilum* (4). At least some of these functions are linked to a ubiquitin and adenosine triphosphate (ATP)-dependent pathway of protein degradation involving the 26S proteasome, an exceptionally large complex ( $M_r \sim 2.0$  MD) (5). The 20S proteasome, or multicatalytic protease complex, forms the catalytic core of the 26S complex, whereas 19S "caps," which associate with the termini of this core have regulatory functions (6). The 26S proteasome differs in its enzymatic properties from the 20S particle. The ATP-dependent degradation of ubiquitinated and of certain nonubiquitinated proteins (7) is

linked to the 26S complex, whereas the 20S proteasome only degrades unfolded proteins in an energy-independent manner (8). Electron microscopic studies have shown that the quaternary structure of the 20S complex is conserved from *Thermoplasma* to humans (9). The cylindrical or barrel-shaped bodies, with dimensions of approximately 150 Å in height and 110 Å in diameter, show a characteristic pattern of four stacked rings of subunits perpendicular to the cylinder axis. The proteasome of the archaeobacterium *T. acidophilum* contains only two different subunits,  $\alpha$  and  $\beta$  (25.8 kD and 22.3 kD, respectively), whereas proteasomes from eukaryotes have a more complex subunit pattern, consisting of multiple  $\alpha$  and  $\beta$  type subunits related to either  $\alpha$  or  $\beta$  subunits of the *Thermoplasma* proteasome. The *Thermoplasma* proteasome has therefore been referred to as an "urproteasome" (10). The  $\alpha$  and  $\beta$  subunits of the *Thermoplasma* proteasome exhibit sequence homology, indicating that they evolved from a common ancestral gene. The hallmark of the  $\alpha$ -type subunits is an  $\text{NH}_2$ -terminal extension of about 35 amino acid residues not present in the  $\beta$ -type subunits. In spite of numerous efforts to characterize the nature of the protease activity, its mechanism has remained unclear. On the basis of studies with small synthetic peptides and protease inhibitors, up to five distinct proteolytic activities have been assigned to the eukaryotic proteasome (11).

The aforementioned studies have not allowed unequivocal classification of the enzyme as a serine, aspartate or thiol protease, and sequence analysis of proteasome subunits has not revealed similarities with

any other known protease (10). Site-directed mutagenesis experiments with recombinant *Thermoplasma* proteasomes exclude the possibility that this proteasome is an unusual cysteine- or serine-type protease (12) as suggested by inhibitor studies. Our structural studies described below and the site directed mutagenesis experiments, described by Seemüller *et al.* (13) show that the proteasome represents a novel type of protease. Degradation studies performed with the *Thermoplasma* proteasome have shown that the substrate proteins are cleaved rather unspecifically. Nevertheless, the degradation products fall into a narrow size range, with most of the fragments having from six to nine amino acid residues, suggesting that some kind of molecular ruler exists (14). This is of particular relevance in view of the role eukaryotic proteasomes have in antigen presentation (15).

**Structure determination.** We have determined the crystal structure from the 20S proteasome of *T. acidophilum* at 3.4 Å resolution. The recombinant protein was purified from *Escherichia coli* as described by Zwickl *et al.* (16) and crystallized as described (17). Briefly, a solution of proteasomes (7 mg/ml) in 5 mM 2-morpholino-propane-sulfonic acid (MPS), pH 7.5, 1 mM EDTA, and 1 mM  $\text{NaN}_3$  was mixed (3:1) with a solution of 10 percent polyethylene glycol (PEG)-1000, 0.1 M potassium phosphate buffer, pH 7.5, and equilibrated against 15 percent PEG-1000, 0.1 M potassium phosphate, pH 7.5. Crystals grew within 1 week to a final size of 0.8 by 0.5 by 0.2 mm<sup>3</sup> at room temperature. Fresh crystals diffract to about 3.2 Å resolution, but their diffraction quality decreases rapidly in the x-ray beam. The crystals belong to the orthorhombic space group  $P2_12_12_1$  and contain one proteasome in the asymmetric unit. The unit cell constants are  $a = 311.9$  Å,  $b = 209.0$  Å, and  $c = 117.2$  Å (Table 1). In that the proteasomes showed sevenfold symmetry when observed by electron microscopy (18), we expected sevenfold noncrystallographic symmetry. This was indeed shown by a self-rotation function (17), and in addition seven noncrystallographic dyads were found perpendicular to the sevenfold axis, which was tilted by 6.4° with respect to the crystallographic  $a$  axis. The structure was solved by single isomorphous replacement (SIR) with the use of noncrystallographic symmetry. The large number of subunits and heavy atom binding sites in the asymmetric unit made the localization of the heavy atom positions difficult. In addition, most crystals showed significant nonisomorphism which was not indicated by cell constant changes but became evident by comparison of diffraction intensities. As an initial test of the usefulness of derivatives, a self-rotation function

J. Löwe, D. Stock, and R. Huber are at the Max-Planck-Institut für Biochemie, Abteilung für Strukturforschung, B. Jap, P. Zwickl, and W. Baumeister are at the Max-Planck-Institut für Biochemie, Abteilung für Strukturbiochemie, Am Klopferspitz 18a, D-82152 Martinsried, Germany.

\*Present address: Life Sciences Division, Lawrence Berkeley Laboratory, University of California, Berkeley, CA 94720, USA.

†To whom correspondence should be addressed.

was calculated for the difference Patterson maps and analyzed for sevenfold symmetry. We found some promising compounds, including  $\text{Ta}_6\text{Br}_{14}$  (TABR) (19), Pt-terpyridinium-chloride (20), and  $\text{Cs}[\text{OsCl}_3(\text{CO})_3]$  (OSCO), but further interpretation by vector-verification methods was not successful. Another derivative,  $\text{Na}_9\text{HSiW}_9\text{O}_{34}$  (W9) (21), showed only two sites in the analysis of the Harker sections, but was not sufficient for structure determination. To overcome the problem of nonisomorphism, we collected native and derivative data sets at 7 Å resolution by soaking the same crystal in the TABR solution and measuring it again. The self-rotation function of the resulting difference Patterson map showed a very clear sevenfold symmetry.

The positioning of the local symmetry-

related heavy atoms was achieved by consecutive rotational and translational searches. Using the self-vectors, we searched for a heptagonal set of heavy atom sites centered at the origin and oriented as revealed by the self-rotation function. We then made a translational search in the cross-vectors for that heptagonal set. Each translational search was performed in two dimensions on the Harker sections, which gave consistent coordinates for the heptagon of heavy atoms and determined the local symmetry operator. Vector verification with this local symmetry operator revealed no further heavy atom positions. The seven sites were refined independently without anomalous differences. The initial SIR-phased electron density map at 7 Å was not interpretable, but difference Fourier maps of two other

derivative data sets, W9 and OSCO, clearly showed heavy atom sites. Those derivatives were used to determine the handedness of the TABR phases, but did not improve the electron density and were omitted. We assumed that the heptagon of TABR sites represented the central plane of the proteasome, thus satisfying the 72-point symmetry and constructed a cylindrical envelope of 154 Å in length and 115 Å in diameter, the largest possible dimension without overlaps from neighboring molecules. The local symmetry axis was refined by a grid search, wherein we varied the orientation or translation. This grid search was followed by three steps of cyclic averaging about the sevenfold axis on a 4 Å grid monitoring the twofold symmetry correlation in the resulting electron density map. After convergence, the averaged map was used to calculate a new envelope by summing density in spheres of 20 Å in diameter with an appropriate cut-off. Starting at 7 Å resolution, 26 steps of averaging around the sevenfold axis and consecutive phase extension to 4.5 Å resolution were calculated on a 2 Å grid. At this stage, a further refinement step of the local symmetry axis was performed as described above with a 1 Å grid and phases at 4.5 Å resolution. The resulting parameters of the sevenfold local symmetry axis were sufficient to extend phases from 4.5 to 3.5 Å in 23 steps. The resulting twofold symmetry had a correlation of 80 percent of the identity operation, and the electron density map was averaged 14-fold for final phase improvement. The resulting electron density was of very high quality, and the complete model could easily be fitted except for the  $\text{NH}_2$ -terminal 12 amino acids of the  $\alpha$  subunit, which are not visible in the electron density (Fig. 1 and Table 1).

**Structures of the subunits.** The core of the  $\alpha$  subunit is a sandwich of two five-stranded antiparallel  $\beta$  sheets with topology S8, S1, S2, S9, and S10 in the upper  $\beta$  sheet and S7, S6, S5, S4, and S3 in the lower  $\beta$  sheet, respectively (Fig. 2A). These have a left-handed twist of about 50° and 80° between the terminal strands. The strand directions of the two  $\beta$  sheets enclose an angle of about 30°. The  $\beta$  sandwich is open at one side, where it is decorated with the  $\text{NH}_2$ -terminal  $\alpha$  helix H0, which fills the cleft. The opposite side of the  $\beta$  sandwich is closed by four hairpin loops connecting  $\beta$  strands S1-S2, S9-S10, S5-S6, and S3-S4, respectively. The upper and lower  $\beta$  sheets of the sandwich are linked by an extended loop between S2 and S3, which forms a main contact site to the neighboring  $\alpha$  subunit via H3 and S8 and the loop between S7 and S8. The  $\beta$  sandwich is flanked by the  $\alpha$  helices H3, H4, and H5 on top and by H1 and H2 at the bottom, which are antiparallel and tilted by 45° relative to



Fig. 1. Stereo view of the original SIR phased electron density at 3.5 Å resolution, 14-fold averaged. The map is contoured at 1 $\sigma$  above background and shows residues 123 to 143 of the  $\beta$  subunit. Superimposed is the current model at 3.4 Å resolution after refinement.

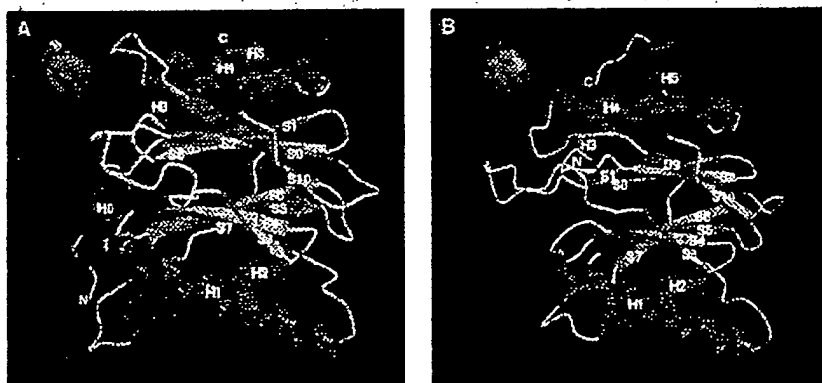


Fig. 2. Ribbon drawing of (A) the  $\alpha$  subunit;  $\alpha$  helices are labeled by H,  $\beta$  strands by S. The  $\text{NH}_2$ -terminal helix H0 represents the front side of the whole proteasome. The identically oriented subunit within the complex is indicated (shown as a yellow sphere) at the top left corner; each subunit is represented by a sphere around its center of gravity. (B) the  $\beta$  subunit in an according orientation (drawings made with MOLSCRIPT (48) and RASTER3D (49).

## RESEARCH ARTICLE

each other. Residues Met<sup>1</sup> to Ile<sup>12</sup> were not visible in the electron density and were therefore omitted from the model. Residues 180 to 208 have weaker than average electron density and the detailed backbone conformation may be uncertain in some places. Proteasomal  $\alpha$  subunits contain a putative nuclear localization sequence (NLS) and the complementary NLS (22), which have been proposed to interact with each other in a phosphorylation-dependent manner. These signals are conserved in the *Thermoplasma* proteasome (23). The NLS sequence comprising residues 49 to 56 is situated at the end of S2, and the following extended loop and is accessible from the solvent. This is also true for a putative complementary NLS (cNLS), which is located at residue 201 to 206 in the loop connecting H4 and S9.

In the course of our model building, the structures of the  $\alpha$  and  $\beta$  subunits emerged to be strikingly similar. The C $\alpha$  atoms of the two subunits can be superimposed with a root-mean-square (rms) deviation of 1.33 Å (Fig. 3B). The structure of the  $\beta$  subunit (Fig. 2B) deviates predominantly in its NH<sub>2</sub>-terminal part from the  $\alpha$  subunit, where it lacks residues 1 to 35. The polypeptide chain of the  $\beta$  subunit starts with strand S1. The absence of  $\alpha$  helix HO provides access to the interior of the  $\beta$  sandwich. The other major differences in the folding of the  $\beta$  subunit comprise the extended loop region between S2 and S3, the loop between H2 and S5, where the  $\alpha$  subunit contains an insertion of four residues, and the COOH-terminal helix H5, which is more extended and followed by an additional loop in the  $\beta$  subunit. All 203 residues of the  $\beta$  subunit are well defined in the electron density. The  $\alpha$  and  $\beta$  subunits appear to represent a novel fold, although  $\alpha$  helices covering  $\beta$  sheets are a frequently occurring structural motif.

The association within the  $\alpha$  and  $\beta$  rings is rather close, such that 24 and 17 percent of the surface of the  $\alpha$  and  $\beta$  subunits, respectively, are covered. In particular, H1 from one  $\alpha$  subunit forms a contact to the H2 helix of the neighboring  $\alpha$  subunit. The connecting loop between H2 and S5 protrudes into the channel of the heptagon of  $\alpha$  subunits and contacts the symmetry-related loops of the neighboring  $\alpha$  subunits on both sides, generating a narrow constriction. The loop preceding HO is situated in close proximity to this site and contacts the HO helix of the next  $\alpha$  subunit. A further contact between  $\alpha$  subunits is established by the loop region between S4 and H1 of one subunit to S7 of the next subunit. The  $\beta$  subunits form a heptagon with similar contacting segments as described for the  $\alpha$  subunits except for the contacts established by the HO helix of  $\alpha$ , which is missing in the  $\beta$  subunit. The major contact between  $\alpha$  and  $\beta$  rings is formed by

the H1 and H2  $\alpha$  helices (Fig. 4A). The wedge-shaped H1, H2 substructure penetrates an opening between subunits of the other ring. This interaction is quasi-symmetrical by homologous segments of  $\alpha$  and  $\beta$ . The buried surface is 10 percent and 11 percent of each of the  $\alpha$  and  $\beta$  rings. The dyad related contacts between the  $\beta$  rings cover 14 percent of the surface of one  $\beta$  ring. They comprise mainly dyad-related segments in H3 and H4 and the loops between H4 and S9 and between S2 and S3, which contact their symmetry mate.

**General architecture and inhibitor binding.** The overall shape of the 20S proteasome

is an elongated cylinder having a central penetrating channel with large cavities and narrow constrictions (Fig. 4, B and D). The overall dimensions of the complex are 148 Å in length and 113 Å in diameter. As would be expected from previous immuno-electron microscopy studies (24), the  $\alpha$  subunits are located at the ends, whereas the  $\beta$  subunits form the two equatorial rings. As also revealed by electron tomography (25), the central channel has three large cavities. Two of them are located at the interface between the  $\alpha$  and the  $\beta$  rings, the third is formed by the  $\beta$  rings in the center of the molecule. The maximum diameter of this cavity is 53

**Table 1.** Collection of data and statistics of structure determination and refinement. Initial collection of native data sets and the screening for heavy atom derivatives was done in house at 14°C with a 180-mm Mar research imaging plate detector (41) mounted on a Rigaku rotating anode x-ray generator operated at 4.8 kW ( $\lambda = \text{CuK}\alpha = 1.5418 \text{ \AA}$ ). Graphite-monochromatized x-rays were collimated to 0.5 mm in diameter. The crystal to detector distance was 350 mm to separate the spots in the  $a^*$  direction. Absorption at this distance was reduced with the use of a helium chamber. Diffraction data of the NAT1 and the corresponding TABR derivative crystal were collected to 7 Å resolution. The TABR-crystal was soaked for 1 hour in 1 mM TABR dissolved in 15 percent PEG-1000, 15 percent propan-2-ol, and 0.1 M potassium phosphate, pH 7.5. The final data set to 3.4 Å resolution was collected at the BW6-beamline at DESY/Hamburg ( $\lambda = 1.0 \text{ \AA}$ ) from six different crystals. Diffraction data were collected with a 300-mm Mar research imaging plate at a distance of 430 mm. The total number of measured reflections in the merged data set was 793,206, with 90,045 unique reflections between 10 and 3.4 Å resolution. All x-ray intensities were evaluated with the MOSFLM program package version 5.2 (42). Scaling and merging were done with CCP4 (43), difference Patterson maps were calculated and interpreted with PROTEIN (44), and real space averaging was performed with MAIN (45). The initial model was built at 3.5 Å resolution with FRODO (44) on an ESV-30 Graphic system work station (Evans & Sutherland, Salt Lake City, Utah). The model was completely fitted in the first density and refined by energy and 14-fold noncrystallographic symmetry restrained crystallographic refinement with X-PLOR (45), with parameters as described by Engh and Huber (47). After refinement,  $2F_o - F_c$  Fourier maps were calculated, and the model was checked and rebuilt. The resolution was extended from 3.5 to 3.4 Å during the refinement procedure. Individual B factors were refined with the use of 14-fold noncrystallographic restraints and restraints between bonded atoms. The current model has been refined to a crystallographic R factor of 22.1 percent, with data from 10 to 3.4 Å resolution, with  $I/\sigma_I > 2.0$  (86,922 unique reflections) and comprises 45,892 nonhydrogen atoms maintaining strict geometry with deviations from ideal values as shown in the table. The model comprises residues 13 to 233 of the 233 amino acids of the  $\alpha$  subunit and all the 203 residues of the  $\beta$  subunit.

	NAT1	TABR	W9	OSCO
Sites (No.)		7	2	14
Completeness (%)				
Overall	92.2	93.4	89.2	92.6
7.43–7.0 Å	58.2	63.6	61.9	58.0
$R_{\text{int}}$		0.16	0.20	0.28
Phasing power		1.04	1.11	1.47
Resolution shells (Å)	10–5.82	4.78	4.23	3.87
NAT1 2–7				
Completeness (%)	88.6	92.3	93.6	94.2
$I/\sigma_I$	14.4	15.0	14.0	12.0
$R_{\text{merge}}^{\dagger}$	0.06	0.08	0.08	0.10
Model				
R factor	22.1	19.4	19.6	22.6
rms bond length (Å)		0.01		
rms angles (°)		1.7		
rms NCS related atoms (Å)		0.16		
rms B bonded atoms (Å)		3.6		
rms B NCS related (Å)		3.7		
Average B $\alpha$ subunit (Å)		51.2		
Average B $\beta$ subunit (Å)		35.2		

<sup>†</sup> $R_{\text{merge}} = \sum |F_{\text{calc}} - F_{\text{obs}}| / \sum F_{\text{calc}}$ , where  $F_{\text{calc}}$  and  $F_{\text{obs}}$  are the derivative and the native structure factor amplitudes, respectively. <sup>‡</sup>Phasing power:  $F_{\text{calc}}/\text{residual}$ ; rms mean heavy atom contribution/rms residual, defined as:  $[(F_{\text{calc}} - F_{\text{obs}})/N]^{1/2}$  with the sum over all reflections, where  $F_{\text{calc}}$  is the calculated structure factor and  $F_{\text{obs}}$  is the structure factor amplitude of the heavy atom derivative, respectively. <sup>§</sup> $R_{\text{merge}} = \sum |I_{\text{obs}} - \langle I_{\text{obs}} \rangle| / \sum I_{\text{obs}}$ , where  $I_{\text{obs}}$  is the intensity value of the  $i$ -th measurement of  $h$  and  $\langle I_{\text{obs}} \rangle$  is the corresponding mean value of  $I$  for all  $i$  measurements of  $h$ ; the summation runs over all measurements. <sup>¶</sup>The rms deviation in bond lengths and bond angles from ideal values. <sup>||</sup>The rms deviation between noncrystallographic symmetry related atom positions. <sup>|||</sup>The rms deviation in temperature factors between bonded and noncrystallographic symmetry related atoms.

Å. Access to these compartments is controlled by four narrow gates, which are formed by the loops connecting H2 and S5 in both subunits. In the  $\alpha$  subunit, the loop Tyr<sup>126</sup>, Gly<sup>127</sup>, and Gly<sup>128</sup> protrudes into the channel and forms a hydrophobic ring around the channel, leaving a passage of only 13 Å in diameter at its entrance. The homologous loops of the  $\beta$  subunits form a bottleneck of 22 Å in diameter in the central part of the complex, ringed by Tyr<sup>93</sup> and Met<sup>96</sup> residues, which protrude into the channel (Fig. 5A).

For inhibitor binding studies, a proteasome crystal was soaked for 36 hours with acetyl-Leu-Leu-norleucinal [LLnL, calpain inhibitor I (Bachem)] a potent inhibitor of the proteasome's Suc-Leu-Leu-Val-Tyr-AMC cleaving activity (26). This peptide aldehyde is an inhibitor of cysteine proteases, such as cathepsin B (27). The unaveraged difference electron density at 4 Å resolution shows the tripeptide bound to all 14  $\beta$  subunits. The tripeptide is situated with its aldehyde function at the NH<sub>2</sub>-terminal threonine of the  $\beta$  subunit in a hydrophobic cleft inside the central cavity of the complex (Fig. 5). The three amino acids of the inhibitor (I1 to I3) are in extended conformation, and form hydrogen bonds from I2 to Thr<sup>21</sup> on one side and from I3 to Gly<sup>47</sup> and from I1 to Val<sup>49</sup> on the other side. The bound inhibitor peptide thus extends the short  $\beta$  strand S1 of the  $\beta$  subunit. The active site flanking residues 20, 21, and 22 belong to the extended loop between S2 and S3, and residues 46, 47, and 48 form the loop between S4 and H1. The hydrophobic cleft is formed by Ile<sup>45</sup>, Ala<sup>46</sup>, Leu<sup>48</sup>, and Val<sup>49</sup> on one side and by Val<sup>20</sup>, Thr<sup>21</sup>, and Met<sup>22</sup> on the other side. In addition, residues 112 to 120 of the neighboring  $\beta$  subunit come close to the active site, and Asp<sup>114</sup> may interact with the N-acetyl group of the inhibitor, but this possibility is not resolved by the present electron density data. Residues Lys<sup>33</sup> and Glu<sup>17</sup> form a salt bridge and are in close proximity to Thr<sup>1</sup>. The difference electron density shows no conformational changes in the protein on binding the inhibitor.

In accordance with the lack of sequence similarity to other proteins, the  $\alpha$  and  $\beta$  subunits appear to exhibit a novel topology. The folding of the two subunits is strikingly similar, despite the insignificant amino acid sequence similarity (Fig. 3A). Differences are seen at the NH<sub>2</sub>- and COOH-termini, notably the NH<sub>2</sub>-terminal  $\alpha$  helix HO of the  $\alpha$  subunit, and in loop regions (in particular at the S2-S3 and the H2-S7 loop). Eukaryotic proteasomal subunits related to the archaeobacterial  $\alpha$  subunit display highly conserved NH<sub>2</sub>-terminal regions (23, 28). This  $\alpha$  helix may be crucial for the interaction with regulatory complexes such as the 19S complex. The helix is situated near the

entrance gate to the channel; the preceding 12 amino acid residues, which have no defined electron density, are disordered so that their precise role is uncertain. The  $\alpha$  subunits assemble spontaneously to form seven-membered rings, whereas  $\beta$  subunits do not. Since deletion of the first 35 amino acid residues in the  $\alpha$  subunit abolishes this ability, the NH<sub>2</sub>-terminal extension appears to have a critical role in assembly (28). The predicted position of HO within the channel (29) is at variance with the observed structure. The four protein disks stack to form tight contacts between the  $\alpha/\beta$  and  $\beta/\beta$  rings, respectively. The  $\alpha/\beta$  contact is established mainly by the  $\alpha$  helices H1 and H2, and the  $\beta/\beta$  contact is formed by  $\alpha$  helices H3, H4, and loops H4-S9 and S2-S3. The  $\beta/\beta$  and  $\alpha/\beta$  double disks have strict and approximate dyad symmetry, respective-

ly. The tight packing appears to make the particle impenetrable from the outer walls of the cylinder (Fig. 4C), allowing access only through the central channel (Fig. 4, B and D). The more constricted outer gates are formed by the loop between H2 and S5 of the  $\alpha$  subunit, which is more extended than in the  $\beta$  subunit and allows the passage of extended polypeptide chains only. The inner gates are formed by equivalent segments of the  $\beta$  subunits. The outer and inner gates control the exchange between solvent and the proteasome and between outer and inner proteasomal cavities, respectively. The conserved tyrosine residues of those loops which protrude into the channel may play a crucial role in structuring the gates and in assisting the passage of substrates and products. The unfolded state of the substrate may be maintained by the hydrophobic en-

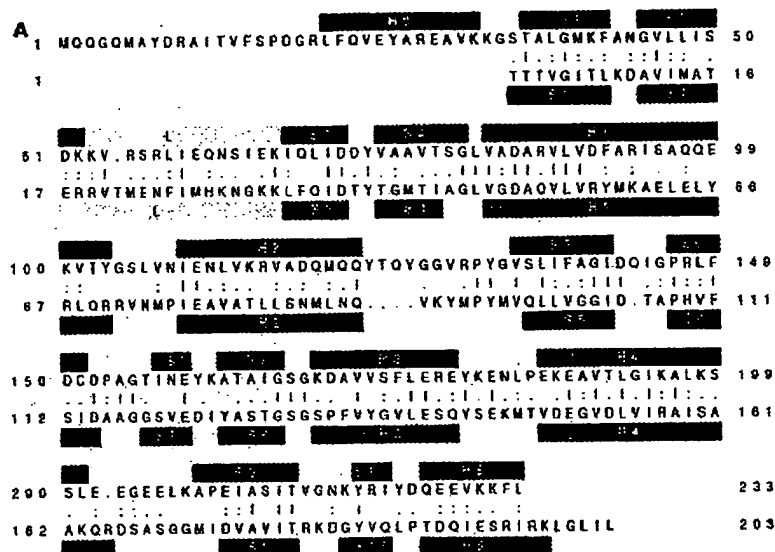


Fig. 3. (A) Sequence alignment of the  $\alpha$  and of the  $\beta$  subunit [BESTFIT, UWGOG package (50)] with a sequence identity of 26 percent. The secondary structure deduced from the refined model, calculated with DSSP (51), is superimposed. Abbreviations for the amino acid residues are A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr. (B) The backbone of the two subunits superimposed with an rms deviation of 1.33 Å with 84 percent of the aligned sequence (GRASP) (52). The rms fit was done with MAIN (43), including residues  $\alpha$ 35- $\alpha$ 52/ $\beta$ 1- $\beta$ 18,  $\alpha$ 66- $\alpha$ 123/ $\beta$ 33- $\beta$ 90,  $\alpha$ 130- $\alpha$ 142/ $\beta$ 93- $\beta$ 105,  $\alpha$ 145- $\alpha$ 202/ $\beta$ 107- $\beta$ 164, and  $\alpha$ 209- $\alpha$ 233/ $\beta$ 172- $\beta$ 196. The orientation of the molecules is the same as in Fig. 2.

## RESEARCH ARTICLE

environment of the channel's interior.

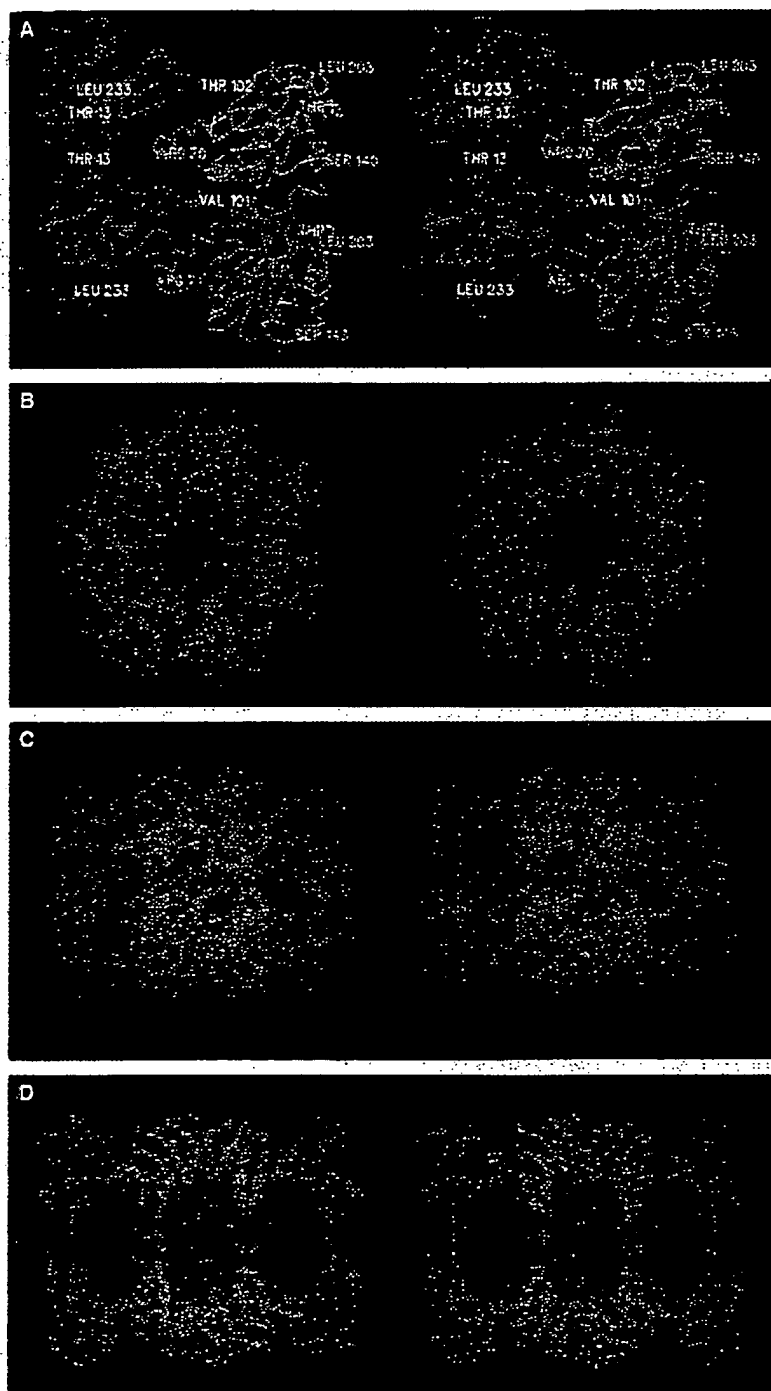
The 20S proteasome has some features in common with the bacterial chaperonin

GroEL (30). The x-ray structure of GroEL (31) allows a more detailed comparison of the two proteins. GroEL is an 850-kD com-

plex consisting of one 14-fold repeating subunit satisfying 72 point group symmetry, and participates in ATP-dependent polypeptide chain folding. Indeed both protein complexes have a central channel through the whole molecule. However, while the interior is divided into three distinct compartments and access to these compartments is controlled by narrow channels in the proteasome, the GroEL channel is more or less continuous. GroEL has relatively large side windows, while the proteasome is tightly sealed. Moreover, there is no similarity in the tertiary structure of the subunits.

Although the LLnL inhibitor model was fitted to a 4 Å resolution map, its position and conformation is well defined by electron density and stereochemical considerations except for the 11 and 12 side chains, which are solvent exposed and may be mobile (Fig. 5C). The inhibitor is bound in a  $\beta$ -sheet conformation between residues 46 to 48 and 20 to 22, which belong to the extended loop between S2 and S3. This loop is stabilized by five main contacts of which four are contacts to symmetry-related  $\beta$  subunits. This may explain why monomeric processed  $\beta$  subunits are not active (28). The aldehyde function of the LLnL inhibitor is situated close to the side chain of Thr<sup>1</sup> of the  $\beta$  subunit, and may form a hemiacetal with the hydroxyl group of Thr<sup>1</sup>. In conjunction with the site directed mutagenesis experiments described by Seemüller *et al.* (13), this suggests that the Thr<sup>1</sup>Oy acts as the nucleophile in peptide hydrolysis. Close to it is Lys<sup>33</sup> N $\epsilon$  and Glu<sup>17</sup> O $\epsilon$ . We propose that together with the NH<sub>2</sub>-terminal amino group, these residues form a catalytic tetrad. In the prototypical serine proteases, the alcoholic function of a Ser residue is activated by a neighboring His, which acts as proton acceptor (32). Analogous functions could be provided in the proteasome by Thr<sup>1</sup>Oy and Lys<sup>33</sup>N $\epsilon$  or Thr<sup>1</sup>N; Glu<sup>17</sup> orients Lys<sup>33</sup> in analogy to the role of Asp<sup>189</sup> in chymotrypsin. As Lys<sup>33</sup> is probably protonated and unlikely to act as proton acceptor, we propose the Thr<sup>1</sup> amino group for this role. Thr<sup>1</sup>Oy-H...N may form a five-membered ring structure in favorable hydrogen bonding geometry. The positively charged Lys<sup>33</sup> could further lower the pK<sub>a</sub> of Thr<sup>1</sup>N by electrostatic effects. The carbonyl groups of Ala<sup>168</sup> and Arg<sup>19</sup> form a hydrogen bond to the NH<sub>2</sub>-terminus, thus orienting and polarizing it. The oxyanion of the tetrahedral transition state, which develops in the acylation reaction may be stabilized by the main chain amide of Gly<sup>47</sup>, which points toward the aldehyde group of the bound inhibitor.

The substrate specificity of the archaeobacterial 20S proteasome is not yet clearly defined, but cleavage seems to be rather unspecific (14). Indeed, the structure of the complex shows a large pocket for the P1

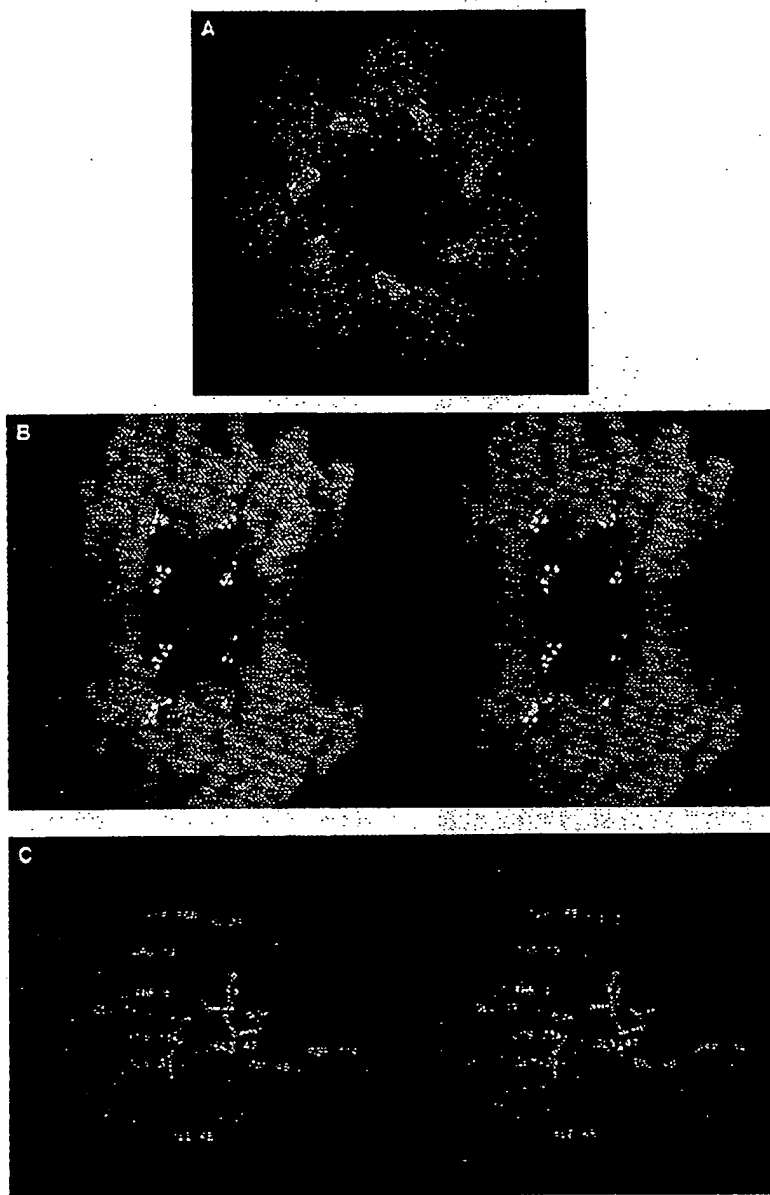


**Fig. 4.** (A)  $\alpha$  drawing of two  $\alpha$  (red) and two  $\beta$  subunits (blue) showing the major contacts between the subunits. (B) Top view of the 20S proteasome showing  $\alpha$  atoms only. (C) Side view of the 20S proteasome. The overall dimensions are 148 Å in length, 113 Å maximum diameter and 75 Å minimum diameter. (D) View of the proteasome cut open along the sevenfold axis; the three compartments and the gates are clearly visible. Dimensions: entrance to the channel:  $\phi$  13 Å, bottlenecks in the center:  $\phi$  22 Å, left and right cavity:  $\phi$  50 Å and length 40 Å, central cavity:  $\phi$  53 Å and length 38 Å.

norleucine, mainly because of Gly<sup>31</sup>. Other residues in contact with the norleucine at P1 are mainly hydrophobic as described above. The S1 pocket is open toward the solvent and may accept side chains of different size and charge. The Leu P2 side

chain is not in direct contact with the protein and is probably unrestricted in size. Our study provides no information on possible binding sites beyond P3, and longer peptides may be relatively free to move. This may be important in explaining the

proteasome's ability to generate peptides of defined length. In eukaryotes, these peptides are exported to the cell surface and presented by MHC class I molecules to initiate the T cell immune response (33). The *Thermoplasma* proteasome also shows a distribution of peptide products centered at hepta- or octapeptides. In line with previous considerations, we suggest that this is achieved by the spatial distribution of the active sites (14). If the acylenzyme intermediate generated in the initial proteolytic reaction has sufficient lifetime, the bound substrate might pass through the central cavity until it meets 1 of the other 13 symmetry related active sites. The shortest distance between two active sites is 28 Å, which may be spanned by a hepta- or octapeptide in extended conformation. Longer peptides may be generated by two active sites at a wider distance, shorter ones could be due to premature deacylation, rebinding, and processing. A relatively stable acylenzyme intermediate could also explain the complete sequential degradation of proteins by eukaryotic proteasomes (34) as the residual polypeptide chain remains bound to an active site of the proteasome. Our model also predicts an oligo(ucra)peptidyl carboxypeptidase activity. It is obvious that the inhibitor peptide cannot be continued in regular conformation in the P' direction (Fig. 5, A and B); therefore, it may turn away from the protein where it would have freedom to move or it may have binding sites on the protein surface. Data with suitable analog inhibitors should provide answers to these possibilities. These structural features explain the lack of specificity of the proteolytic action of the proteasome. The implication of lysine and hydroxyl groups is not without precedent and has been reported for the leader peptidase from *E. coli* (35). Lysines and serines in the active sites of serine proteases have been classified by Rawlings and Barrett (36) and shown to occur in a family containing the serine-type D-Ala-D-Ala carboxypeptidase and  $\beta$  lactamase by crystal structures. The latter enzyme opens  $\beta$ -lactam rings, which mimic peptide bonds. The active site of  $\beta$  lactamase from *Staphylococcus aureus* consisting of Ser<sup>70</sup>, Lys<sup>73</sup>, and Glu<sup>166</sup> (37) shows a geometry similar to that of Thr<sup>1</sup>, Lys<sup>33</sup>, and Glu<sup>17</sup> in the proteasome although in a totally different structural context. The D-Ala-D-Ala carboxypeptidase catalyzes the crosslinking of peptidoglycans in bacterial cell walls by transpeptidation. These transpeptidases contain lysine, serine, and glutamate in their active sites (38). The transpeptidase reaction requires a long-lived acylenzyme intermediate (39), possibly also important for the proteasome. The structure of a penicillin acylase suggests the NH<sub>2</sub>-terminal serine residue as the catalytic



**Fig. 5.** (A)  $F_o - F_c$  difference Fourier map of the Ac-Leu-Leu-norleucinal inhibitor (LLnL) at 4 Å resolution. Superimposed is the current model of the  $\beta$  subunit. Fourteenfold averaged difference electron density showing seven inhibitors sitting around the central cavity formed by  $\beta$  subunits. Distances between active sites within one  $\beta$  ring are 28 Å, 51 Å, and 64 Å. (B) Stereo drawing of the accessible surface of the central cavity built by two rings of  $\beta$  subunits cut open along the sevenfold axis. Superimposed is a space-filling model of the LLnL inhibitors (GRASP) (52). (C) Stereo drawing of the active site with a 14-fold averaged  $F_o - F_c$  map. Superimposed is the nonrefined inhibitor model (MAIN) (45). Data collection of the inhibitor data set: a proteasome crystal was soaked in 5 mM LLnL dissolved in 15 percent PEG-1000, 15 percent propan-2-ol, 0.1 M potassium phosphate, pH 7.5, for 36 hours. A complete data set to 4 Å resolution was collected and evaluated as described in Table 1 for NAT1 with a 300-mm imaging plate [Mar research (41)].



## RESEARCH ARTICLE

nucleophile similar to the  $\text{NH}_2$ -terminal threonine of the proteasome (40).

The *Thermoplasma* 20S proteasome closely resembles in size, shape, and functional properties the eukaryotic 20S proteasome. All subunits of the eukaryotic proteasomes are related by amino acid sequence similarity to the  $\alpha$  or  $\beta$  subunit of the *T. acidophilum* proteasome, suggesting a very similar catalytic mechanism (13). The much simpler subunit composition and internal symmetry of the archaeobacterial proteasome was the key to structure determination, which now provides a detailed model for the elucidation of the structure and biochemical roles of eukaryotic proteasomes.

## REFERENCES AND NOTES

1. J. M. Peters, *Trends Biochem. Sci.* **19**, 377 (1994); A. J. Rivett, *Biochem. J.* **291**, 1 (1993); A. L. Goldberg, *Eur. J. Biochem.* **203**, 9 (1992); M. Rechsteiner, L. Hoffman, W. Dubiel, *J. Biol. Chem.* **268**, 60 (1993); A. Hershko and A. Ciechanover, *Annu. Rev. Biochem.* **61**, 761 (1992).
2. V. J. Palombella, O. J. Rando, A. L. Goldberg, T. Maniatis, *Cell* **78**, 773 (1994); A. Ciechanover, *ibid.* **79**, 13 (1994); B. Richter-Ruoff and D. H. Wolf, *FEBS Lett.* **338**, 34 (1993).
3. K. L. Rock et al., *Cell* **78**, 761 (1994).
4. B. Dahlmann et al., *FEBS Lett.* **251**, 125 (1989).
5. J. M. Peters, Z. Cijka, J. R. Harris, J. A. Kleinschmidt, W. Baumeister, *J. Mol. Biol.* **234**, 932 (1993).
6. J. M. Peters, W. W. Franko, J. A. Kleinschmidt, *J. Biol. Chem.* **269**, 7709 (1994); L. Hoffman, G. Pratt, M. Rechsteiner, *ibid.* **267**, 22362 (1992); T. Yoshimura et al., *J. Struct. Biol.* **111**, 200 (1993); G. N. DeMartino et al., *J. Biol. Chem.* **268**, 20887 (1994).
7. Y. Murakami et al., *Nature* **360**, 597 (1992).
8. J. M. Fagan and L. Waxman, *J. Biol. Chem.* **267**, 23015 (1992); T. Wenzel and W. Baumeister, *FEBS Lett.* **326**, 215 (1993); W. Matthews, K. Tanaka, J. Driscoll, A. Ichihara, A. L. Goldberg, *Proc. Natl. Acad. Sci. U.S.A.* **86**, 2579 (1989).
9. G. Puchler, F. Pitzer, P. Zwickl, W. Baumeister, *Syst. Appl. Microbiol.* **16**, 734 (1994).
10. P. Zwickl et al., *Biochemistry* **31**, 964 (1992); K. Tanaka, T. Tanura, T. Yoshimura, A. Ichihara, *Now Biol.* **4**, 173 (1992); W. Heinemeyer, N. Tröndle, G. Albrecht, D. H. Wolf, *Biochemistry* **33**, 12229 (1994).
11. M. Orlowski, C. Cardozo, C. Michaud, *Biochemistry* **32**, 1563 (1993); H. Djaballah and A. J. Rivett, *ibid.* **31**, 4133 (1992).
12. E. Seemüller, A. Lupas, F. Zühl, P. Zwickl, W. Baumeister, *FEBS Lett.* **359**, 173 (1995).
13. E. Seemüller et al., *Science* **268**, 579 (1995).
14. T. Wenzel, C. Eckerskorn, F. Lottspeich, W. Baumeister, *FEBS Lett.* **349**, 205 (1994).
15. M. G. Brown, J. Driscoll, J. J. Morawo, *Nature* **353**, 355 (1991); R. Glynn et al., *ibid.*, p. 357; J. Driscoll and D. Finley, *Cell* **68**, 823 (1992); A. L. Goldberg and K. L. Rock, *Nature* **357**, 375 (1992).
16. P. Zwickl, F. Lottspeich, W. Baumeister, *FEBS Lett.* **312**, 157 (1992).
17. B. Jap et al., *J. Mol. Biol.* **234**, 681 (1993).
18. G. Puchler et al., *EMBO J.* **11**, 1607 (1992).
19. P. H. Kuhn and R. E. McCarley, *Inorg. Chem.* **4**, 1482 (1965).
20. H. M. Brothers II and N. M. Kostio, *Biochemistry* **29**, 7168 (1990).
21. R. G. Finke, B. Rapko, R. J. Jaxton, P. J. Domaille, *J. Am. Chem. Soc.* **108**, 2947 (1986).
22. K. Tanaka et al., *FEBS Lett.* **271**, 41 (1990).
23. P. Zwickl, F. Lottspeich, B. Dahlmann, W. Baumeister, *ibid.* **278**, 217 (1991).
24. A. Grziwa, W. Baumeister, B. Dahlmann, F. Kopp, *ibid.* **290**, 186 (1991).
25. R. Hegert, G. Pfeifer, G. Pühler, B. Dahlmann, W. Baumeister, *ibid.* **283**, 117 (1991).
26. B. Dahlmann, L. Kuelin, A. Grziwa, P. Zwickl, W. Baumeister, *Eur. J. Biochem.* **208**, 189 (1992).
27. T. Sasaki et al., *J. Enzyme Inhibition* **3**, 165 (1990).
28. P. Zwickl, J. Klein, W. Baumeister, *Nature Struct. Biol.* **1**, 765 (1994).
29. A. Lupas, A. J. Koster, T. Wutz, W. Baumeister, *FEBS Lett.* **354**, 45 (1994).
30. P. Zwickl et al., *J. Struct. Biol.* **103**, 197 (1990).
31. K. Braig et al., *Nature* **371**, 578 (1994).
32. R. Huber and W. Bodo, *Acc. Chem. Res.* **11**, 114 (1978).
33. L. R. Dick et al., *J. Immunol.* **152**, 3884 (1994).
34. L. R. Dick, C. R. Moomaw, G. N. DeMartino, C. A. Staughter, *Biochemistry* **30**, 2725 (1991).
35. W. R. Tschantz, M. Sung, V. M. Delgado-Partin, R. E. Dalbey, *J. Biol. Chem.* **268**, 27349 (1993).
36. N. D. Rawlings and A. J. Barrett, *Biochem. J.* **290**, 205 (1993).
37. C. C. H. Chen, J. Rahil, R. F. Pratt, O. Herzberg, *J. Mol. Biol.* **234**, 165 (1993).
38. J. A. Kelly, J. R. Knox, H. Zhao, J. Frère, J. Ghuyen, *ibid.* **209**, 261 (1989).
39. J.-M. Ghuyen et al., *Top. Mol. Pharmacol.* **1**, 63 (1981).
40. H. J. Dugglesby et al., *Nature* **373**, 264 (1995).
41. Z. Dauter, H. Terry, H. Witzel, K. S. Wilson, *Acta Crystallogr.* **B46**, 833 (1990).
42. A. G. W. Leslie, Joint CCP4 and ESRF-EACMB Newsletter, Protein Crystallogr. 26 (Daresbury Laboratory, Warrington, UK, 1992).
43. CCP4, Science and Engineering Research Council (SERC), Collaborative Computing Project no. 4 (SERC Daresbury Laboratory, Warrington, UK, 1979).
44. W. Steigemann, *Crystallographic Computing* **5**, D. Moras, A. D. Podjarny, J. C. Thiery, Eds. (Oxford Univ. Press, Oxford, 1991).
45. D. Turk, Ph.D. thesis, Technische Universität München (1992).
46. T. A. Jones, *J. Appl. Crystallogr.* **11**, 268 (1978).
47. A. T. Bruenger, X-PLOR, Version 3.1: A System for Crystallography and NMR (Yale Univ. Press, New Haven, CT, 1992); R. A. Engh and R. Huber, *Acta Crystallogr.* **A47**, 392 (1991).
48. P. J. Kraulis, *J. Appl. Crystallogr.* **24**, 946 (1991).
49. E. A. Merritt and M. E. P. Murphy, *Acta Crystallogr.* **D50**, 869 (1994).
50. Genetics Computer Group, Wisconsin, USA (1991).
51. W. Kabsch and C. Sander, *Biochemistry* **22**, 2577 (1983).
52. A. Nicholls and B. Honig, Columbia University (1993).
53. We thank M. Stubbs for critical reading of the manuscript and H. D. Bartnik for help with the x-ray data collection. Supported in part by Fonds der Chemischen Industrie (D.S.). The coordinates have been deposited in the Protein Data Bank, Brookhaven, NY, with tracking number 15750 for the coordinates and 15751 for the observed structure factor amplitudes.

20 December 1994; accepted 10 March 1995

## AAAS-Newcomb Cleveland Prize

To Be Awarded for a Report, Research Article, or an Article Published in *Science*

The AAAS-Newcomb Cleveland Prize is awarded to the author of an outstanding paper published in *Science*. The value of the prize is \$5000; the winner also receives a bronze medal. The current competition period began with the 3 June 1994 issue and ends with the issue of 26 May 1995.

Reports, Research Articles, and Articles that include original research data, theories, or syntheses and are fundamental contributions to basic knowledge or technical achievements of far-reaching consequence are eligible for consideration for the prize. The paper must be a first-time publication of the author's own work. Reference to pertinent earlier work by the author may be included to give perspective.

Throughout the competition period, readers are invited to nominate papers appearing in the Reports, Research Articles, or Articles sections. Nominations must be typed, and the following information provided: the title of the paper, issue in which it was published, author's name, and a brief statement of justification for nomination. Nominations should be submitted to the AAAS-Newcomb Cleveland Prize, AAAS, Room 924, 1333 H Street, NW, Washington, DC 20005, and must be received on or before 30 June 1995. Final selection will rest with a panel of distinguished scientists appointed by the editor-in-chief of *Science*.

The award will be presented at the 1996 AAAS annual meeting. In cases of multiple authorship, the prize will be divided equally between or among the authors.